

Section 4.0 Baseline Sampling Duration & Frequency

4.1 Power and Sample Size

To determine an adequate sample size (number of samples), EPA will require approximate statistical power of at least 75 percent ($\pi = 0.75$) for a statistical decision procedure used to detect a difference of one standard deviation over a period of one year, while requiring a Type I error rate (significance level) of 5 percent ($\alpha = 0.05$). Lower power is acceptable on a monthly (or single-observation) basis, but cumulative and annual decision procedures are required to have at least 75 percent power over the course of one year.

Power, in the context of this discussion, quantifies the probability that a particular statistical test and sample size will indicate that the mean or median loading has increased over the baseline, given that it truly has increased some specified amount (see Table 4.1a). The test is designed to guard against incorrectly concluding that the mean or median has increased by setting alpha at a low value. The probability is *less than* or equal to alpha that a statistical test and sample size will incorrectly indicate that the mean or median loading has increased over the baseline, given that it has *not* increased. If there has been a decrease in loadings, the risk of such an incorrect decision will be considerably less than alpha.

Power was calculated for the one-sided, two-sample t-test with alpha = 0.05 (Table 4.1a). This does not apply exactly to alternative statistical decision procedures, but EPA believes that it provides a reasonable basis for calculating a minimum sample size. The effect of autocorrelation upon power of the t-test was examined (using an approximate method) and was found to be small.

Table 4.1a: Power of one-sided, two-sample t-test with equal n in each group, $\alpha = 0.05$

N	Difference in means divided by sigma, $(\mu_2 - \mu_1) /$					
	0.5	1.0	1.5	2.0	2.5	3.0
2	0.07	0.10	0.15	0.23	0.36	0.53
3	0.10	0.21	0.39	0.62	0.80	0.90
4	0.13	0.31	0.57	0.79	0.92	0.97
5	0.16	0.39	0.69	0.89	0.97	0.99
6	0.18	0.47	0.77	0.94	0.98	1.00
7	0.21	0.53	0.84	0.96	0.99	1.00
8	0.23	0.59	0.88	0.98	1.00	1.00
9	0.25	0.64	0.91	0.99	1.00	1.00
10	0.27	0.69	0.94	0.99	1.00	1.00
11	0.29	0.73	0.96	1.00	1.00	1.00
12	0.31	0.76	0.97	1.00	1.00	1.00
13	0.33	0.80	0.98	1.00	1.00	1.00
14	0.35	0.82	0.98	1.00	1.00	1.00
15	0.37	0.85	0.99	1.00	1.00	1.00
16	0.39	0.87	0.99	1.00	1.00	1.00
17	0.41	0.88	0.99	1.00	1.00	1.00
18	0.42	0.90	1.00	1.00	1.00	1.00
19	0.44	0.91	1.00	1.00	1.00	1.00
20	0.46	0.93	1.00	1.00	1.00	1.00
21	0.47	0.94	1.00	1.00	1.00	1.00
22	0.49	0.95	1.00	1.00	1.00	1.00
23	0.51	0.95	1.00	1.00	1.00	1.00
24	0.52	0.96	1.00	1.00	1.00	1.00
Values were calculated using function "t.test.power()" from Environmental Stats for S-Plus (Millard, 1998)						

To estimate the Type I and Type II error rates for more complex decision procedures, simulations or resampling of a large baseline dataset would be needed. The Type I and Type II error rates of the statistical procedures proposed in this document have not yet been fully evaluated by EPA.

To approximate the power likely to be provided (in the least favorable case) by a most-powerful *non-parametric* test, EPA note's that the Wilcoxon-Mann-Whitney test has asymptotic relative efficiency (ARE) at least 0.864 compared to the t-test, and may have $ARE > 1$ for heavy-tailed distributions.

The criterion for sample size applied above was the ability to detect a change of one standard deviation above baseline loadings with high power. An increase of one standard deviation can represent a large increase in loading, given the large variability of flows and loadings observed in mine discharges. The coefficient of variation (CV) is the ratio of standard deviation to mean. Sample CVs for iron loadings range approximately from 0.25 to 4.00, and commonly exceed 1. Sample CVs for manganese loadings range approximately from 0.24 to 5.00. When the CV equals 1, an increase of the average loading by one standard deviation above baseline means a doubling of the loading.

Loadings from mine discharges appear to be approximately distributed lognormally. Thus, logarithms of loadings may be approximately distributed normally, justifying use of a t-test.

Based on these considerations and the power of the t-test (Table 4.1a), EPA has determined that the smallest acceptable number and frequency of samples is 12 monthly samples, taken consecutively over the course of one year. This number is approximate and represents the absolute minimum. Twelve samples may provide less than the required power if autocorrelation is very high, if sampling duration is less than a year, or if the sampling interval is shortened (e.g., to one week) while the number of samples is not increased above 12. In such cases, more samples should be taken, enough to provide the required statistical power.

A permitting authority may want to consider the statistical power appropriate for environmental protection, and may decide to require more than twelve samples per year during and after the baseline year in order to increase power and in order to provide a fair chance of observing a representative sample of discharge flows and loadings.

4.2 Sampling Plan

Based on considerations described above, EPA requires a minimal sampling plan for establishing baseline and post-baseline pollution load that consists of taking one sample per month for one year. The duration, frequency, and seasonal distribution of sampling are important aspects of a sampling plan, and can affect the precision and accuracy of statistical estimates as much as can the number of samples.

Sampling, during and after baseline, should systematically cover all periods of the year during which substantial discharge flows can be expected. Sampling should not bias the baseline mean toward high loadings by over-sampling the high-flow months. Unequal sampling of different time periods can be accounted for using statistical estimation procedures appropriate to stratified sampling.

Twelve samples are unlikely to provide a representative sample of discharges for the year, given the variability observed for coal mine discharges. Eighteen to twenty-four samples seem much more likely to adequately characterize a baseline year.

EPA proposes a minimal sampling plan for baseline and post-baseline that consists of taking one sample per month for one year.

There may be acceptable alternatives to the proposed minimum duration and frequency of one sample per month for twelve months. The merits of alternative sampling plans have not been evaluated thoroughly. Alternative plans could be based upon subdivision of the year into distinct

time periods that might be sampled with different intensities, and other types of stratified sampling plans (e.g., Sanders et al., 1983, and Griffiths, 1990) that attempt to account for seasonal variations. Seasonal stratification has the potential to provide a basis for more precise estimates of baseline characteristics, if the sampling plan is designed and executed correctly and if results are calculated using appropriate statistical estimators.

Sampling should be designed to prevent biased sampling. In particular, flow measurement methods should accurately measure flows during high-flow events. If the discharge overflows or bypasses the weir or flume, or if a measurement is not made as scheduled on a high-flow day, statistical characterizations of flow and loading will be inaccurate. The sampling location and methods should be designed as much as possible to permit access and sampling on all scheduled days, and to avoid the need to reschedule sampling because flow is extremely high.

References

- Griffiths, J.C., 1990. Letter to M. Smith, Pennsylvania Department of Environmental Resources dated January 14, 1990, 3 pages, with attachment of 54 pages titled "Chapter 3, Statistical Analysis of Mine Drainage Data".
- Millard, S.P., 1998. Environmental Stats for S-Plus: Users Manual for Windows and Unix. Springer-Verlag: New York, NY.
- Sanders T.G., R.C. Ward, J.C. Loftis, T.D. Steele, D.D. Adrian, and V. Yevjevich, 1983. Design of Networks for Monitoring Water Quality. Water Resources Publications, Littleton, CO.

